

### **Supplementary Information**

## **Novel antibody epitopes dominate the antigenicity of spike glycoprotein in SARS-CoV-2 compared to SARS-CoV**

Ming Zheng\*, Lun Song\*

### **Authors' Affiliations:**

Institute of Military Cognition and Brain Sciences, Academy of Military Medical Sciences, 27 Taiping Road, Beijing 100850, China

\* Corresponding authors.

**Ming Zheng, MD**, Institute of Military Cognition and Brain Sciences, Academy of Military Medical Sciences, 27 Taiping Road, Beijing 100850, China (Telephone: +86-10-66930324, E-mail: mmzheng@fmmu.edu.cn); **Lun Song, PhD**, Institute of Military Cognition and Brain Sciences, Academy of Military Medical Sciences, 27 Taiping Road, Beijing 100850, China (Telephone: +86-10-66930320, E-mail: lunsong0752@163.com).

## **Materials and Methods**

### *Acquisition and processing of sequence data.*

A total of 10 whole genome sequences of SARS-CoV-2 (2019-nCoV) were downloaded on 30 January 2020 from the GISAID database<sup>1</sup>, including EPI\_ISL\_402119, EPI\_ISL\_402120, EPI\_ISL\_402121, EPI\_ISL\_402123, EPI\_ISL\_404228, EPI\_ISL\_403935, EPI\_ISL\_404895, EPI\_ISL\_406597, EPI\_ISL\_406844 and EPI\_ISL\_406862. Those 10 SARS-CoV-2 genomes were collected from different parts in the world (Wuhan, Zhejiang, Guandong in China and other countries of USA, France, Australia and Germany). Next, these nucleotide sequences were aligned to the GenBank reference sequence (accession ID: NC\_045512.2) and then translated into amino acid residues according to the coding sequence positions, and provided along the reference sequence for the SARS-CoV-2 spike proteins. Also, the spike protein sequences for SARS-CoV (accession ID: SARS-GD01: AY278489; SARS-GZ02: AY390556; SARS-NS-1, AY508724; SARS-Sino1-11, AY485277; SARS-Tor2: NC\_004718) and MERS-CoV (NC\_019843) were obtained following the same procedure from GenBank.

### *Constructing the phylogenetic tree of multiple sequence alignment*

Multiple sequence alignment was conducted on the spike protein sequences of SARS-CoV, MERS-CoV, and SARS-CoV-2 using CLUSTAL W<sup>2</sup>. Next, the pairwise scoring matrix of the distances was computed using Fitch similarity from aligned sequences<sup>3</sup>. And, based on the resulting scoring matrix, phylogenetic tree was constructed using neighbor-joining method<sup>4</sup>.

### *Analysis of antibody epitope*

The sequence-based antibody epitopes score was analyzed using BepiPred-2.0 bioinformatic tool<sup>5</sup>, and the antibody epitope candidates were analyzed using the default threshold value of 0.50. The surface epitope accessibility was computed using Emini surface accessibility scale method<sup>6</sup>. Next, the antibody epitope candidates were filtered by the surface accessible epitopes score using the default threshold value of 1.0 (Fig. 1k).

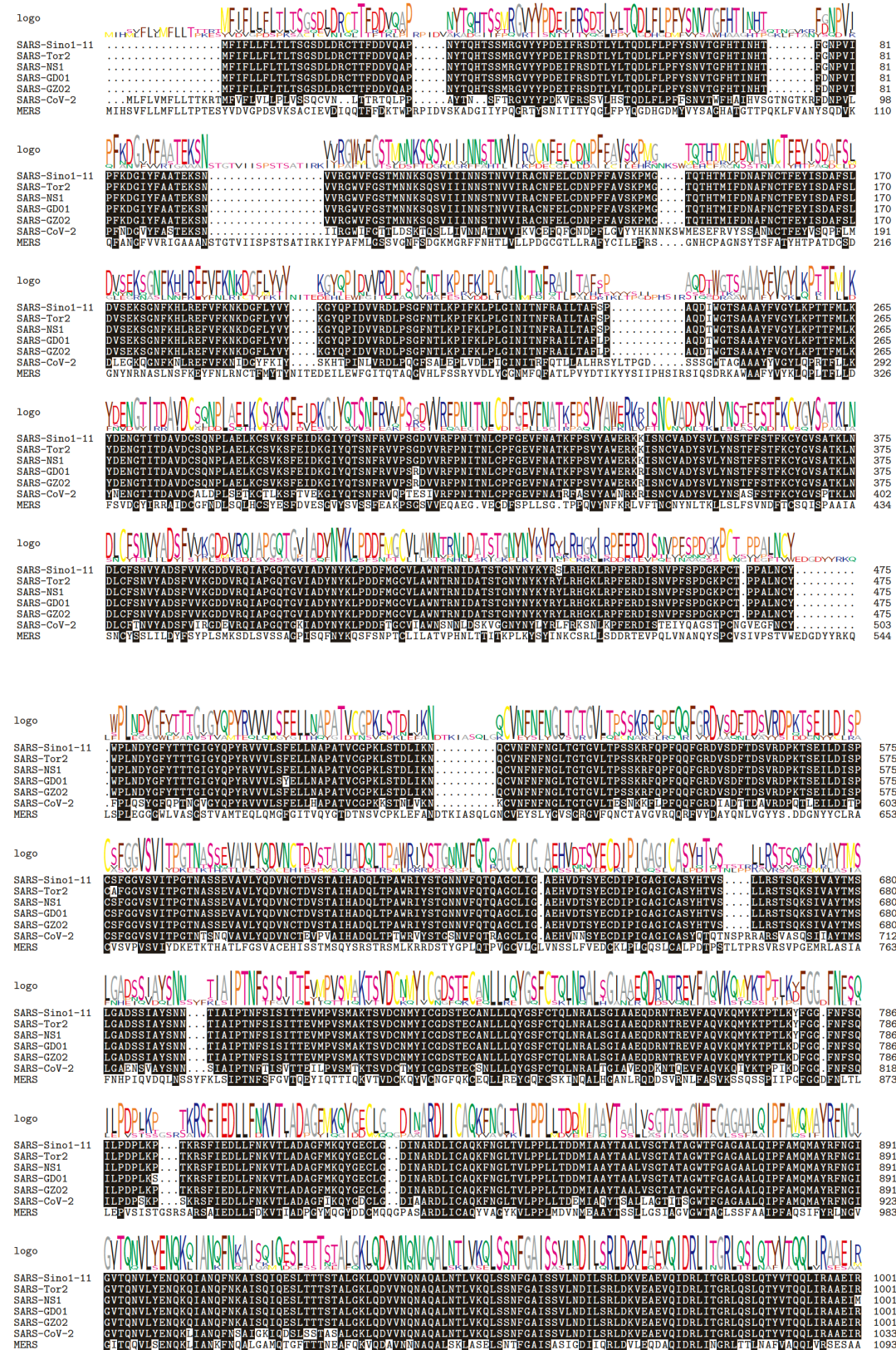
### *Construction of 3-D protein structure model*

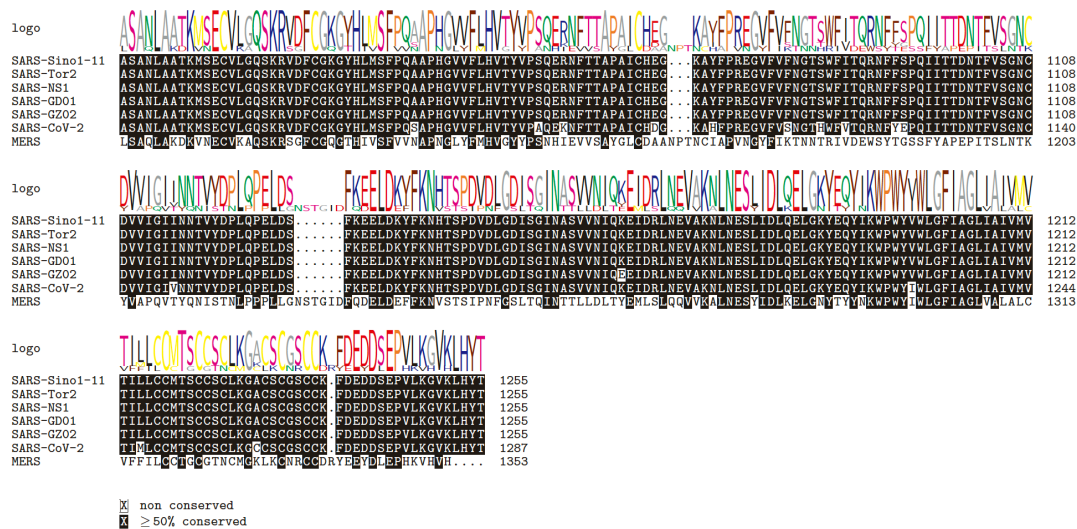
The 3-D cryo-electron microscopy structure of the SARS-CoV spike protein complexed with its host cell receptor ACE2 protein were derived from Protein Data Bank (PDB accession: 6ACJ) <sup>7</sup>. Next, according to the SARS-CoV-2 spike protein sequence, the Swiss-model bioinformatic tool<sup>8</sup> was used to model the 3-D complex structure of the SARS-CoV-2 spike protein binding to ACE2.

### *Statistical Analysis*

All statistical analyses were performed by using R v3.6.1. Wilcoxon signed-rank for two samples and Kruskal–Wallis test with Dunn’s multiple comparison test for multiple samples were performed using the stats package as appropriate. All statistical tests used 0.05 as the significance level, and  $p \geq 0.05$  were considered as not significant (n.s.), and  $p < 0.05$  were considered as significant difference, indicated with the asterisks (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\*  $p < 0.001$  and \*\*\*\*  $p < 0.0001$ ).

## Supplementary Figure 1





**Fig. S1.** Protein sequence alignment of spike proteins in MERS-CoV, SARS-CoV and SARS-CoV-2.



**Fig. S2.** Protein sequence Alignment of the spike proteins in SARS-CoV-2 (2019-nCoV) from Wuhan, Zhejiang, Guandong in China and other countries of USA, France, Australia and Germany showing the conservancy of spike proteins.



### Supplementary Figure 3

Antibody Epitope	Source	a.a. length	Epitope Score	Surface Accessibility	High Score	RBD	Region
FDEDDS	SARS-CoV-2	6	0.532	3.098	✗ No	✗ No	✓ Conserved
IQKEIDRL	SARS-CoV-2	8	0.542	1.665	✗ No	✗ No	✓ Conserved
KYFKNHTSP	SARS-CoV-2	9	0.556	2.464	✓ Yes	✗ No	✓ Conserved
MAYR	SARS-CoV-2	4	0.545	1.200	✗ No	✗ No	✓ Conserved
NVLYENQ	SARS-CoV-2	7	0.563	1.784	✗ No	✗ No	✓ Conserved
QSKR	SARS-CoV-2	4	0.540	2.021	✗ No	✗ No	✓ Conserved
YQPY	SARS-CoV-2	4	0.511	2.063	✗ No	✓ Yes	✓ Conserved
SEFR	SARS-CoV-2	4	0.559	1.366	✗ No	✗ No	✗ Non-conserved
TPGDSS	SARS-CoV-2	6	0.578	1.292	✗ No	✗ No	✗ Non-conserved
TTKR	SARS-CoV-2	4	0.557	2.210	✓ Yes	✗ No	✗ Non-conserved
YYHKNNKSWM	SARS-CoV-2	10	0.570	2.938	✓ Yes	✗ No	✗ Non-conserved
ASTEK	SARS-CoV-2	5	0.577	2.164	✓ Yes	✗ No	★ Conserved+Non-conserved
AWNRRK	SARS-CoV-2	6	0.553	2.597	✓ Yes	✓ Yes	★ Conserved+Non-conserved
DPSKPSKRSP	SARS-CoV-2	10	0.517	3.239	✗ No	✗ No	★ Conserved+Non-conserved
DQLTPTWRVY	SARS-CoV-2	10	0.538	1.356	✗ No	✗ No	★ Conserved+Non-conserved
EQDKNTQ	SARS-CoV-2	7	0.578	3.874	✓ Yes	✗ No	★ Conserved+Non-conserved
ESNKK	SARS-CoV-2	5	0.546	2.952	✗ No	✗ No	★ Conserved+Non-conserved
FPQSA	SARS-CoV-2	5	0.538	1.340	✗ No	✗ No	★ Conserved+Non-conserved
GFQPT	SARS-CoV-2	5	0.532	1.315	✗ No	✓ Yes	★ Conserved+Non-conserved
GTNTSN	SARS-CoV-2	6	0.617	1.928	✓ Yes	✗ No	★ Conserved+Non-conserved
HVNNSY	SARS-CoV-2	6	0.593	1.450	✗ No	✗ No	★ Conserved+Non-conserved
IADTTDAVRDPQT	SARS-CoV-2	13	0.543	1.623	✗ No	✗ No	★ Conserved+Non-conserved
IYSKHT	SARS-CoV-2	6	0.512	1.920	✗ No	✗ No	★ Conserved+Non-conserved
KYNENGT	SARS-CoV-2	7	0.564	2.004	✓ Yes	✗ No	★ Conserved+Non-conserved
LDSKTQ	SARS-CoV-2	6	0.617	2.240	✓ Yes	✗ No	★ Conserved+Non-conserved
LKPFERDI	SARS-CoV-2	8	0.524	1.918	✗ No	✓ Yes	★ Conserved+Non-conserved
LTTRTQLPPAYTNS	SARS-CoV-2	14	0.545	1.798	✗ No	✗ No	★ Conserved+Non-conserved
NSNNLD	SARS-CoV-2	6	0.612	1.685	✗ No	✓ Yes	★ Conserved+Non-conserved
PKKS	SARS-CoV-2	4	0.567	2.980	✓ Yes	✓ Yes	★ Conserved+Non-conserved
QTSNFRVQPT	SARS-CoV-2	10	0.559	1.667	✗ No	✓ Yes	★ Conserved+Non-conserved
SMTKT	SARS-CoV-2	5	0.608	1.403	✗ No	✗ No	★ Conserved+Non-conserved
TNGTKRFD	SARS-CoV-2	8	0.537	2.355	✗ No	✗ No	★ Conserved+Non-conserved
VPAQEKNF	SARS-CoV-2	9	0.538	2.113	✗ No	✗ No	★ Conserved+Non-conserved
YQTQTNSPRRAR	SARS-CoV-2	12	0.587	3.348	✓ Yes	✗ No	★ Conserved+Non-conserved

**Fig. S3.** Antibody epitopes of spike proteins from SARS-CoV-2.

## REFERENCES

- 1 Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* **22**, doi:10.2807/1560-7917.ES.2017.22.13.30494 (2017).
- 2 Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-4680, doi:10.1093/nar/22.22.4673 (1994).
- 3 Fitch, W. M. An improved method of testing for evolutionary homology. *Journal of molecular biology* **16**, 9-16, doi:10.1016/s0022-2836(66)80258-9 (1966).
- 4 Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406-425, doi:10.1093/oxfordjournals.molbev.a040454 (1987).
- 5 Jespersen, M. C., Peters, B., Nielsen, M. & Marcatili, P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res* **45**, W24-W29, doi:10.1093/nar/gkx346 (2017).
- 6 Emini, E. A., Hughes, J. V., Perlow, D. S. & Boger, J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol* **55**, 836-839 (1985).
- 7 Song, W., Gui, M., Wang, X. & Xiang, Y. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathog* **14**, e1007236, doi:10.1371/journal.ppat.1007236 (2018).
- 8 Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**, W296-W303, doi:10.1093/nar/gky427 (2018).